A Modified TLS-Prony Method Using Data Decimation

William M. Steedly, *Member, IEEE,* Ching-Hui J. Ying, *Student Member, IEEE,* and Randolph L. Moses, *Senior Member, IEEE*

***Abstract*—This paper introduces a modified TLS-Prony method that incorporates data decimation. The use of data decimation results in the reduction in the computational complexity because one high-order estimation is replaced by several low-order estimations. We present an analysis of pole variance statistics for this modified TLS-Prony method. This analysis provides a quantitative comparison of the parameter estimation accuracy as a function of decimation factors. We show that by using decimation, one can obtain comparable statistical performance results at a fraction of the computational cost, when compared with the conventional TLS-Prony algorithm.**

I. Introduction

A popular high-resolution estimation technique is the use of backward linear prediction coupled with singular value decomposition (SVD) and total least squares [1], which, here, is called the TLS-Prony technique. This technique has been shown to provide good parameter estimates of damped exponential signals in noise for various types of data [1], [2]. However, for large data lengths, the TLS-Prony method can be computationally expensive. The reason for this is that the TLS-Prony method involves computing the singular value decomposition (SVD) of a data matrix of size (m, n), where m is related to data length and n to prediction order. For best accuracy of the parameter estimates, $n \approx \frac{m}{3}$. Thus, this data matrix becomes quite large for longer data lengths. To overcome this problem, it is sometimes possible to decimate the data before applying the TLS-Prony technique; the result is often a large reduction in computations. In this paper, we consider the statistical and computational properties of the TLS-Prony algorithm when used in conjunction with data decimation.

Data decimation has been considered before in the context of spectral estimation [3], [4]. This technique entails using only part of the measured data. Decimation of correlation sequences was also considered in [5]; this technique effectively uses all the measured data but is somewhat restrictive in that it applies only to correlation-based parameter estimation techniques. These works do not present a quantitative analysis of statistical properties of the resulting parameter estimates.

Manuscript received September 9, 1991; revised November 25, 1993. This work was supported by the Air Force Office of Scientific Research, Bolling AFB, DC, and the Avionics Division, Wright Laboratories, Wright Patterson AFB, OH.

W. M. Steedly is with The Analytic Sciences Corporation (TASC), Reston, VA 22090 USA.

C. J. Ying and R. L. Moses are with the Department of Electrical Engineering, The Ohio State University, Columbus, OH 43210 USA.

IEEE Log Number 9403288.

In this paper, we develop a data decimation technique based on the TLS-Prony algorithm [1]. We also present a theoretical statistical analysis of the accuracy of the TLS-Prony parameter estimates when decimation and any linear FIR filter are used. Based on this analysis, we present a quantitative comparison of estimation accuracy for various types of data decimation schemes. In particular, we compare the decimated and nondecimated procedures in terms of estimation accuracy. Our analysis demonstrates that the performance of the decimated TLS-Prony procedure is comparable to the performance of the nondecimated TLS-Prony procedure for undamped exponential modes. We also develop a complexity analysis and show that the decimated algorithm is computationally more efficient than the nondecimated algorithm. Both the statistical performance and the decrease in computational complexity are verified by Monte-Carlo simulations.

In particular, we apply the statistical analysis to consider two specific cases of interest. First, the signals of interest may be bandlimited and occupy a relatively small region of the unambiguous frequency range $f \in \left[-\frac{1}{2}, \frac{1}{2}\right]$. In this case, one is interested in analyzing a subset of the whole frequency range. For example, this technique was used to investigate radar signatures of aircraft [2], [6]. We consider whether or not decimation can improve the performance of estimation in this case.

A second case of interest is when the signal occupies most or all of the unambiguous frequency range. In this case, we filter the data to isolate a number of subbands and then use a decimation version of TLS-Prony to estimate the poles in each of the subbands. This idea is similar in principle to beamspace prefiltering in array processing [7]–[11]. By focusing on particular bands one at a time, estimation techniques can be used with lower model orders since there are typically fewer modes within each of the bands. Thus, a single wideband estimation procedure is replaced by several lower order estimations. This has the advantages of being much less numerically intensive and of being amenable to parallel implementation [10].

An outline of this paper is as follows. In Section II, we develop the modified TLS-Prony procedure. In Section III, we derive the first-order approximation of the statistics of the estimated parameters. In Section IV, we develop a procedure and a complexity analysis for performing full spectrum estimation. We also discuss filter design and performance loss in the estimation. Section V presents some simulation studies using decimation. Finally, Section VI concludes the paper.

II. Decimation Estimation Procedure

A. Data Model

Assume we have N "snapshots" of data vectors $y(t)$, each of length m:

$$y(t) = [y_0(t)\ y_1(t)\ \cdots\ y_{m-1}(t)]^T \quad t = 1, 2, \ldots, N. \tag{1}$$

Each data vector is modeled as a noisy exponential sequence

$$y_q(t) = \sum_{i=1}^{n} x_i(t) p_i^q + e_q(t) \quad q = 0, 1, \ldots, m-1. \tag{2}$$

There are n distinct exponential modes in the data; the n poles $\{p_i\}_{i=1}^n$ do not vary from snapshot to snapshot, but the amplitudes $x_i(t)$ may vary. Here, it is assumed that $\{e_q(t)\}$ are uncorrelated zero mean complex white Gaussian noise sequences with variance σ. Equation (2) may be compactly written as

$$y(t) = Ax(t) + e(t), \tag{3}$$

where $e(t) = [e_0(t)\ \ e_1(t)\ \ \cdots\ \ e_{m-1}(t)]^T$, $x(t) = [x_0(t)\ \ x_1(t)\ \ \cdots\ \ x_{n-1}(t)]^T$, and A is the $m \times n$ Vandermonde matrix derived from n signal poles

$$A = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ p_1 & p_2 & \cdots & p_n \\ p_1^2 & p_2^2 & \cdots & p_n^2 \\ \vdots & \vdots & & \vdots \\ p_1^{m-1} & p_2^{m-1} & \cdots & p_n^{m-1} \end{bmatrix}. \tag{4}$$

B. Parameter Estimation

Consider the $m \times 1$ data vector $y(t)$ as given by (1). In general, we first filter the data set $y(t)$ before decimating to minimize effects of aliasing (we discuss filtering in detail in Section IV). Here, we consider an lth-order FIR filter of the form

$$y'(t) = Ky(t) \tag{5}$$

where

$$K = \begin{bmatrix} k_l & k_{l-1} & k_{l-2} & \cdots & k_0 & 0 & 0 & \cdots & 0 \\ 0 & k_l & k_{l-1} & \cdots & k_1 & k_0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & k_l & k_{l-1} & \cdots & k_1 & k_0 & 0 \\ 0 & \cdots & 0 & 0 & k_l & \cdots & k_2 & k_1 & k_0 \end{bmatrix}_{(m-l)\times m} \tag{6}$$

and where the sequence $\{k_c\}_{c=0}^l$ is the FIR filter impulse response. The resulting filtered sequence is of length $m - l$ and does not include what would be the first l transient points. This filter plays the same role as beamformers used for array processing problems to isolate particular beamwidths of interest (e.g., see [9], [11]).

From $y'_q(t)$, we now define a set of decimated sequences as

$$y'^u_q(t) = y'_{qd+u}(t) \qquad \begin{matrix} q = 0, 1, \ldots, m'_d - 1 \\ u = 0, 1, \ldots, d-1 \end{matrix} \tag{7}$$

where $m'_d = \lfloor \frac{m-l}{d} \rfloor$. The index "$u$" gives the start sample in the decimation; thus, the sequences $\{y'^u_q(t)\}$ (for fixed t) represent a set of interleaved sequences decimated from $y'(t)$. These sequences are sometimes referred to as "polyphase components" (e.g., see [12]). From (2) and (7), we see that each sequence $\{y'^u_q(t)\}$ is a noisy exponential sequence of the form

$$y'^u_q(t) = \sum_{i=1}^{n} x'^u_i(t)(p'_i)^q + e''^u_q(t) \tag{8}$$

where

$$\begin{aligned} p'_i &= (p_i)^d \\ x'^u_i(t) &= x_i(t) p_i^{(u+l)} \mathcal{K}(p_i) \end{aligned} \tag{9}$$

and where $\mathcal{K}(z)$ is the FIR filter polynomial given by

$$\mathcal{K}(z) = k_0 + k_1 z^{-1} + k_2 z^{-2} + \cdots + k_l z^{-l}. \tag{10}$$

Note that $\{e''^u_q(t)\}$ is the corresponding filtered and decimated noise sequence.

The effects of the FIR filtering in the new model are to scale the amplitude coefficients and to color the noise. In general, we will choose K to be a bandpass filter. By careful choice of the FIR filter, we can significantly reduce the mode amplitude coefficients outside of some band of interest; in this case, we can assume the number of the "significant" modes in the filtered data is n', which is less than n. In this case, we have the following model:

$$y'^u_q(t) = \sum_{i=1}^{n'} x'^u_i(t)(p'_i)^q + e'^u_q(t) \tag{11}$$

where

$$e'^u_q(t) = \sum_{i=n'+1}^{n} x'^u_i(t)(p'_i)^q + e''^u_q(t). \tag{12}$$

The noise $e'^u_q(t)$ in (11) contains both the noise $e''^u_q(t)$ and the $n - n'$ residual mode signals; thus, it is colored Gaussian noise and has nonzero mean since the $n - n'$ modes outside the band of interest are still present even though they are attenuated. The effect of the nonzero mean is to introduce some bias in the parameter estimates, as we will see in the following sections.

The extra interleaved data sets can be discarded to reduce computational burden in parameter estimation, i.e., u in (7) can take on only the value 0. From a Nyquist theory point of view, discarding the interleaved data should provide no loss in performance if an ideal low-pass filter can be implemented. However, for real applications, we will lose some performance due to the use of nonideal filters. We can keep all the interleaved data to compensate the nonideal filtering. Nevertheless, the performance improvement is not significant. Since our primary concern is computational efficiency, we will thus focus on keeping only one of the interleaved data sets in the paper. Note that discarding of all but one data set can be easily incorporated into the matrix K by keeping only every dth row.

We now have N decimated multisnapshot sequences. As a result, the TLS-Prony algorithm can be applied to the data

in (8) to give estimates $\{\widehat{p_i'}\}$ and $\{\widehat{x_i'^0}(t)\}$. We thus have decimated multisnapshot backward linear prediction equations given by

$$[y'^0 \quad Y'^0]\begin{bmatrix}1\\ b'\end{bmatrix} \approx 0 \tag{13}$$

where

$$b' = [b_1' \quad b_2' \quad \cdots \quad b_L']^T \tag{14}$$

and (15), which appears at the bottom of the page Here, L is the order of prediction, and b' is the coefficient vector of the polynomial $B'(z)$ given by

$$B'(z) = 1 + b_1'z + b_2'z^2 + \cdots + b_L'z^L. \tag{16}$$

The choice of L affects the accuracy of the coefficients b_i', as we address in later sections.

The TLS-Prony method considers the effect of noise perturbation of both Y'^0 and y'^0, and the TLS solution attempts to minimize the effect of these perturbations on the prediction coefficient vector b' (see [1] for details). This is accomplished by obtaining an SVD of the matrix $[y'^0 : Y'^0]$ and truncating all but the first n' singular values to arrive at an estimate $[\widehat{y'^0} : \widehat{Y'^0}]$ [1]. This TLS procedure ignores some structure in the noise elements, and this results in suboptimal statistical performance (as compared with the Cramér-Rao bound (CRB)). In order to take advantage of the noise structure in the data matrix, one can use a constrained TLS solution technique such as the one in [13]. However, the constrained techniques are computationally intensive. Furthermore, the unconstrained TLS technique gives nearly optimal statistical performance with proper selection of algorithm parameters (see Section V) at substantially reduced computational cost. Since a primary interest in the paper is computational reduction, we proceed using the basic TLS procedure.

Inserting $[\widehat{y'^0} : \widehat{Y'^0}]$ in (13) gives the modified linear prediction equation

$$\widehat{Y'^0}\widehat{b'} \approx -\widehat{y'^0} \tag{17}$$

from which the linear prediction coefficient vector estimate $\widehat{b'}$ is found as

$$\widehat{b'} = -\widehat{Y'^0}^{+}\widehat{y'^0} \tag{18}$$

where $^+$ denotes the Moore–Penrose pseudoinverse. Finally, the estimates for the decimated poles are found by

$$\widehat{p_j'} = \text{zero}_j(\widehat{B'}(z)), \qquad j = 1, 2, \ldots, L. \tag{19}$$

It is not, in general, possible to recover $\widehat{p_j}$ from $\widehat{p_j'}$ because the mapping $p_i \rightarrow p_i^d$ is not one to one. However, the mapping can be made one to one by suitably restricting the domain of p_i. For example, if it is known *a priori* that $\angle p_i \in \left(-\frac{\pi}{d}, \frac{\pi}{d}\right)$, then $\widehat{p_j}$ may be uniquely recovered from $\widehat{p_j'}$ by

$$\widehat{p_j} = \left(\widehat{p_j'}\right)^{\frac{1}{d}}. \tag{20}$$

In order to meet the domain restriction requirement without *a priori* information, one can choose a suitable FIR filter K, as discussed in Section IV.

Once the poles are found, the corresponding amplitude coefficients can be estimated from the decimated pole estimates and decimated data using (11) and (9). Using the decimated pole estimates and decimated data is more computationally efficient since there will be shorter data lengths and fewer pole estimates. For this case, (11) leads to the following equation for the amplitude coefficients:

$$\widehat{A'}_L\widehat{\mathcal{K}_p}\widehat{X} = Y_a'^0, \tag{21}$$

$$[y'^0 \quad Y'^0] = \begin{bmatrix}
y_0'(1) & y_d'(1) & y_{2d}'(1) & \cdots & y_{Ld}'(1)\\
y_d'(1) & y_{2d}'(1) & y_{3d}'(1) & \cdots & y_{(L+1)d}'(1)\\
\vdots & \vdots & \vdots & & \vdots\\
y_{m_d'-(L+1)d}'(1) & y_{m_d'-Ld}'(1) & y_{m_d'-(L-1)d}'(1) & \cdots & y_{m_d'-d}'(1)\\
y_0'(2) & y_d'(2) & y_{2d}'(2) & \cdots & y_{Ld}'(2)\\
y_d'(2) & y_{2d}'(2) & y_{3d}'(2) & \cdots & y_{(L+1)d}'(2)\\
\vdots & \vdots & \vdots & & \vdots\\
y_{m_d'-(L+1)d}'(2) & y_{m_d'-Ld}'(2) & y_{m_d'-(L-1)d}'(2) & \cdots & y_{m_d'-d}'(2)\\
 & & \vdots & & \\
y_0'(N) & y_d'(N) & y_{2d}'(N) & \cdots & y_{Ld}'(N)\\
y_d'(N) & y_{2d}'(N) & y_{3d}'(N) & \cdots & y_{(L+1)d}'(N)\\
\vdots & \vdots & \vdots & & \vdots\\
y_{m_d'-(L+1)d}'(N) & y_{m_d'-Ld}'(N) & y_{m_d'-(L-1)d}'(N) & \cdots & y_{m_d'-d}'(N)
\end{bmatrix} \tag{15}$$

where

$$\widehat{A'}_L = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \widehat{p'_1} & \widehat{p'_2} & \cdots & \widehat{p'_L} \\ \widehat{p'_1}^2 & \widehat{p'_2}^2 & \cdots & \widehat{p'_L}^2 \\ \vdots & \vdots & & \vdots \\ \widehat{p'_1}^{(m'_d-1)} & \widehat{p'_2}^{(m'_d-1)} & \cdots & \widehat{p'_L}^{(m'_d-1)} \end{bmatrix}$$

$$\widehat{\mathcal{K}_p} = \operatorname{diag}\left(\widehat{p_1}^l \mathcal{K}(\widehat{p_1}), \widehat{p_2}^l \mathcal{K}(\widehat{p_2}), \ldots, \widehat{p_L}^l \mathcal{K}(\widehat{p_L})\right)$$

$$\widehat{X} = \begin{bmatrix} \widehat{x}_1(1) & \widehat{x}_1(2) & \cdots & \widehat{x}_1(N) \\ \widehat{x}_2(1) & \widehat{x}_2(2) & \cdots & \widehat{x}_2(N) \\ \vdots & \vdots & & \vdots \\ \widehat{x}_L(1) & \widehat{x}_L(2) & \cdots & \widehat{x}_L(N) \end{bmatrix}$$

$$Y'^0_a = [y'^0(1) \quad y'^0(2) \quad \cdots \quad y'^0(N)]. \tag{22}$$

The amplitude coefficients can be found from a least squares solution to (21)

$$\widehat{X} = \widehat{\mathcal{K}_p}^{-1}\left(\widehat{A'}^*_L \widehat{A'}_L\right)^{-1} \widehat{A'}^*_L Y'^0_a, \tag{23}$$

where * denotes complex conjugate transpose. We note that (23) is not used in practice to solve (22) because more numerically sound procedures (such as QR decomposition [14]) can be used; however, this equation is useful for the statistical analysis presented below.

Because only n' singular values of $\widehat{Y'^0}$ are nonzero, there are, at most, n' pole estimates that can correspond to true data modes. Therefore, only the n' poles that have the largest energy are retained [15]. We then reestimate the amplitude coefficients of these n' poles by eliminating all but the n' "high-energy pole" columns of $\widehat{A'}_L$ and then recomputing the least squares solution for $\widehat{X}$. We note that the second amplitude coefficient estimation can be done by using the QR decomposition from the first amplitude coefficient estimation. By doing so, we save computation in the second amplitude coefficient estimation. Because the noise $e'^u_q(t)$ in (11) is not in general white, an unweighted least solution to (21) may not lead to amplitude coefficient estimates with smallest variance.

III. Statistical Analysis

A major contribution of this paper is the derivation of the statistical properties of the TLS-Prony pole estimates obtained using decimation. Below, we derive a general expression for the first order approximation of the probability distribution function (pdf) of the estimates of $\{p'_i\}$ under the assumptions that there is a filter as described by (5). This expression applies to different decimation values; therefore, it can be used to determine the relative statistical accuracy for various choices in the TLS-Prony algorithm. The expression is given in the following theorem.

Theorem 4.1: Assume we are given FIR filtered data $\{y'^0_q(t)\}$ as defined in (7) and (11). Let

$$\widehat{P'} = \begin{bmatrix} \widehat{p'_1} & \widehat{p'_2} & \cdots & \widehat{p'_{n'}} \end{bmatrix}^T \tag{24}$$

be the n' highest energy TLS-Prony pole estimates found from (19) and (23). Then, the first-order approximation (as $\sigma \to 0$) of the pdf of $\widehat{P'}$ is given by

$$\widehat{P'} \sim \mathcal{N}(P' + P'_b, \Sigma') \tag{25}$$

where the estimate bias and covariance are given by

$$\begin{aligned} P'_b &= F'G'S'^{0+}P'^0_\epsilon, \\ \Sigma' &= \sigma F'G'S'^{0+}\Sigma^0_\epsilon\left(F'G'S'^{0+}\right)^*. \end{aligned} \tag{26}$$

Here, P'^0_ϵ is defined in (62), and Σ^0_ϵ is a block diagonal matrix given by

$$\begin{aligned} \Sigma^0_\epsilon = \operatorname{diag}\Big(&B'K^0K^{0*}B'^*, B'K^0K^{0*}B'^*, \\ &\ldots, B'K^0K^{0*}B'^*\Big)_{N(m'_d-L)\times N(m'_d-L)}. \end{aligned} \tag{27}$$

The expressions for K^0 and B' are given by

$$K^0 = \begin{bmatrix} K(1)^T & K(1+d)^T & \cdots & K(1+m'-d)^T \end{bmatrix}^T \tag{28}$$

where $K(a)$ is the ath row of K, and (29), which appears at the bottom of the page. We also have

$$F' = \operatorname{diag}\left(\frac{1}{\eta'_1}, \frac{1}{\eta'_2}, \ldots, \frac{1}{\eta'_{n'}}\right), \tag{30}$$

$$\eta'_i = \begin{bmatrix} b'_1 & b'_2 & \cdots & b'_L \end{bmatrix} \begin{bmatrix} 1 \\ 2p'_i \\ \vdots \\ Lp'^{(L-1)}_i \end{bmatrix}, \tag{31}$$

$$G' = \begin{bmatrix} p'_1 & p'^2_1 & \cdots & p'^L_1 \\ p'_2 & p'^2_2 & \cdots & p'^L_2 \\ \vdots & \vdots & & \vdots \\ p_{n'} & p'^2_{n'} & \cdots & p'^L_{n'} \end{bmatrix}, \tag{32}$$

and S'^0 defined as the noise-free version of Y'^0.

$$B' = \begin{bmatrix} 1 & b'_1 & b'_2 & \cdots & b'_L & 0 & 0 & \cdots & 0 \\ 0 & 1 & b'_1 & \cdots & b'_{L-1} & b'_L & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & b'_1 & \cdots & b'_{L-1} & b'_L & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & b'_{L-2} & b'_{L-1} & b'_L \end{bmatrix}_{(m'_d-L)\times(m'_d)}. \tag{29}$$

Proof: See the Appendix. □

Equation (26) provides the biases and covariances for decimated pole estimates given a particular set of poles and decimation factor. If the nondecimated pole estimates $\{\hat{p}_j\}$ are recovered from the decimated pole estimates $\{\hat{p}'_j\}$ using (20), the biases and variances for the nondecimated poles can be derived in terms of the biases and variances for the decimated poles using a first-order approximation of (20). Defining $\tilde{p}'_i = \hat{p}'_i - p'_i$, we have the following derivation:

$$\begin{aligned} p'_i + \tilde{p}'_i &= (\hat{p}_i)^d \\ &= (p_i + \tilde{p}_i)^d \\ &= p_i^d + dp_i^{d-1}\tilde{p}_i + \\ &\quad \text{(higher order terms)} \end{aligned} \tag{33}$$

and note that $p'_i = p_i^d$. Thus

$$\tilde{p}_i \approx \frac{\tilde{p}'_i}{dp_i^{d-1}}. \tag{34}$$

Therefore, we have

$$\begin{aligned} \text{Bias}(\hat{p}_i) &\triangleq E[\hat{p}_i] - p_i = \frac{\text{Bias}(\hat{p}'_i)}{dp_i^{d-1}} \\ \text{Var}(\hat{p}_i) &\triangleq E[(\hat{p}_i - E[\hat{p}_i])(\hat{p}_i - E[\hat{p}_i])^*] \\ &= \frac{\text{Var}(\hat{p}'_i)}{d^2|p_i|^{2(d-1)}}. \end{aligned} \tag{35}$$

The variances of the nondecimated poles can now be compared with their respective CRB's [16], provided that the estimated bias is negligible. The CRB results in [16] can be directly compared with the variances of the estimated poles using the TLS-Prony method to examine its performance in both nondecimated and decimated circumstances. This comparison is shown in Section V for a number of examples.

IV. Full Spectrum Estimation Using Filtering and Decimation

Using the decimation scheme that has been developed, any poles or modes not in the band of interest need to be filtered out so that they are not aliased into the band of interest by the decimation operation. Even if there are no poles outside the band of interest, a filter can be still applied to reduce the out-of-band noise that will be aliased into the band of interest by the decimation operation. However, the imperfections of an FIR filter, nonideal stopband rejection, and data length reduction by transient response effects will cause loss in performance, as shown in Section V. In general, nonideal stopband rejection increases bias in the estimation (because the leakage of the stopband poles, and thus the first term in (12), will be larger), and data length reduction due to transient effects of the FIR filter causes a variance increase.

In this section, we develop a procedure to obtain full spectrum mode estimates by use of bandwidth segmentation. This is similar to beamspace prefiltering in array processing [7]–[11]. We then present a design procedure for the needed FIR filters. Finally, we present operation counts for both the decimated and nondecimated estimation procedures.

A. Procedure

To examine the use of filtering and decimation, assume we are interested in estimating modes in the complete spectrum (that is, poles that may lie anywhere in the complex plane), and we wish to use a decimation factor of d. We estimate poles in each region of Fig. 1(a) using a decimation based TLS-Prony procedure. First, we modulate the data to center the band of interest about $f = 0$ as follows:

$$y_q(t) \rightarrow y_q(t)e^{-j2\pi f_0 q} \tag{36}$$

where f_0 is the modulating frequency for one of the subbands of interest $f_0 = 0, \frac{1}{d}, \frac{2}{d}, \ldots, \frac{d-1}{d}$. We then low-pass filter the modulated data to isolate the frequency band $f \in \left[-\frac{1}{2d}, \frac{1}{2d}\right]$ (this region corresponds to A and B in Fig. 1(a) for $d = 6$). Finally, we apply the decimated TLS-Prony algorithm of Section II. The resulting pole estimates $\hat{p}'_j$, as given by (20), lie in $f \in \left[-\frac{1}{2}, \frac{1}{2}\right]$ as shown in Fig. 1(b). The corresponding pole estimates in nondecimated frequency space are given by (20) with modulation as follows:

$$\hat{p}_j = \left(\hat{p}'_j\right)^{\frac{1}{d}} e^{j2\pi f_0}. \tag{37}$$

A problem that can result from the above procedure is that poles near the endpoints of the subband region may be incorrectly estimated near the opposite endpoint. This results from the discontinuity of the mapping in (37) for $\angle\hat{p}'_j \approx \pm\pi$. It can be seen from Fig. 1(b) that small errors in estimates near A and B result in large differences in Fig. 1(a). To avoid this problem, we use $c > d$ overlapped estimations, each of size $\frac{1}{d}$ (note that this changes the modulating frequencies to be $f_0 = 0, \frac{1}{c}, \frac{2}{c}, \ldots, \frac{c-1}{c}$). This is shown in Fig. 1(c) for $c = 2d$. For each region, we retain only those pole estimates that are in the half of the overlap region that is closer to the center of the band of interest. For example, in Fig. 1(c) for the subband [A, B], we retain poles only in the region [C, D]. This corresponds to retaining pole estimates whose angles satisfy

$$\angle\hat{p}'_j \in \left[-\frac{\pi d}{c}, \frac{\pi d}{c}\right] \tag{38}$$

in the decimated frequency space. In this way, we reduce the effects of the discontinuity of the mapping in (20) for $\angle\hat{p}'_j$ near π. The overlap method also helps to provide immunity to effects of a nonideal low-pass filter, as is discussed in the next subsection.

B. Filter Design and Performance Loss

As discussed before, the use of finite-length FIR filters results in performance loss. We wish to design a filter such that this performance loss is minimized. Because the stopband of the filter is set by the decimation factor d to be $\left[-\frac{1}{2d}, \frac{1}{2d}\right]$, there remain three free design parameters: the order of the FIR filter (l), the number of overlapped estimations (c), and the flatness of the passband. Note that the passband is defined as $\left[-\frac{1}{2c}, \frac{1}{2c}\right]$, which is the region in which we actually retain the pole estimates.

Each of the parameters has its own effect in the estimation performance. Larger filter lengths result in a variance increase

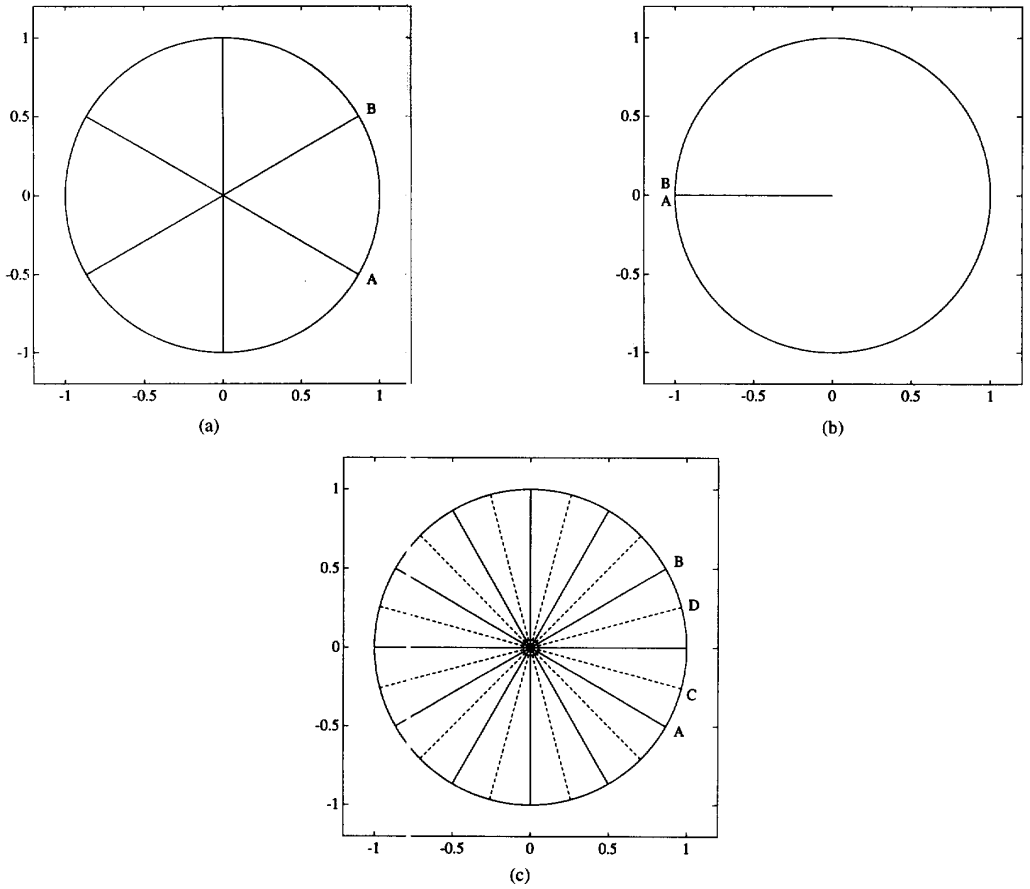


Fig. 1. Spectral effect due to decimating by six.

because of the data length reduction associated with discarding the transient response of the filter output. Shorter filter lengths result in increased bias of the pole estimates due to aliasing of imperfectly attenuated modes in the stopband region. The number of overlapped estimations primarily influences the computational aspect. A larger c results in more computations; nevertheless, a larger c allows for a larger filter transition band, thus allowing us to design for a better filter stopband. Note that we can tolerate some nonflatness since we can compensate for this effect in the amplitude coefficient estimation. However, a nonflat passband increases the variance of the poles near the minima of the passband since the filter reduces the power of these poles more than other poles in the passband.

After filtering and decimating the data, the stop-band signal poles are aliased into the band of interest. If the energy of the filtered stop-band signal modes is small relative to the in-band Gaussian noise, the estimation bias caused by the stop-band signal poles will be negligible in comparison with the estimation variance. This implies that for high SNR signals, we need more stop-band rejection in order to avoid the bias problem. One should choose a filter length sufficient to attenuate out-of-band modes to be below the noise floor.

With the above constraints in mind, we design the filters as follows. First, based on SNR, we determine the needed stop-band rejection to ensure that the leakage of the out-of-band signal poles is small relative to the in-band noise. Then, we choose the filter order l and the number of the overlapped estimations c to give the desired stop-band rejection with a small transition band. Note that we want l and c to be as small as possible. For instance, in the example given below, we estimated according to the SNR that a 20 dB stop-band rejection results in acceptably low bias. We find that an equiripple FIR filter using $c = 2d = 12$ and $l = 20$ gives the approximated stop-band rejection. An equiripple filter is used since pass-band flatness is less important than transition bandwidth because we can compensate for pass-band ripple during the amplitude coefficient estimation. Although this procedure is somewhat *ad hoc*, it seems reasonable and works well in the example below and in other experiments investigated by the authors. Other filter designs have been developed (e.g., see [17]–[19]), and the filter design problem continues to be an active research area.

C. Operation Counts

The main goal of applying the decimation procedure is a reduction in computations. In this section, we compare the computations of a full TLS-Prony estimation applied to a

nondecimated sequence to those of TLS-Prony estimations applied to a set of band-pass filtered and decimated sequences. We include the computations associated with overlapping as well as those associated with the filtering operation for this comparison.

In the operation counts that follow, we assume that only one of the interleaved sets of decimated data is used in the decimation-based TLS-Prony algorithm (as in (13)). In addition, we assume $L = \frac{m}{3d}$ is used for the prediction order because this prediction order gives near optimal accuracy (see [15], [20], and Section V). We compute operation accounts for the SVD operation, the QR decomposition, the polynomial root finding operation, and the filtering operation; these were found to account for over 90% of the total operations in our computer simulations.

SVD Operation Counts: For a real matrix that has dimension $R \times C$, the approximate floating point operation (flop) count associated with the "economical" SVD computation (in which only the first C left singular vectors are computed) is given by $\mathrm{fc}^{\mathrm{SVD}} \approx 14RC^2 + 8C^3$ [14]. For a complex matrix, the count is about a factor of 2 larger. In our case, the matrix $\left[\widehat{y}'\ \widehat{Y}'\right]$ has dimension $R \times C = \left(\frac{2(m-l)}{3d}N\right) \times \left(\frac{(m-l)}{3d} + 1\right)$. Therefore, for the nondecimated case ($d = 1$ and $l = 0$), we get an SVD flop count of

$$\mathrm{fc}^{\mathrm{SVD}}_{\mathrm{nondec}} \approx 28\left(\frac{2}{3}mN\right)\left(\frac{m}{3}+1\right)^2 + 16\left(\frac{m}{3}+1\right)^3 \quad (39)$$

and for the decimated case with c overlap regions, we obtain

$$\mathrm{fc}^{\mathrm{SVD}}_{\mathrm{dec}} \approx c\left(28\left(\frac{2(m-l)}{3d}N\right)\left(\frac{m-l}{3d}+1\right)^2 + 16\left(\frac{m-l}{3d}+1\right)^3\right). \quad (40)$$

QR Decomposition Operation Counts: For a real matrix, the approximate flop count associated with the QR decomposition is given by $\mathrm{fc}^{\mathrm{QR}} \approx 2RC^2 - \frac{2}{3}C^3$ [14]. For complex matrices, the count is a factor of 4 larger. In our case, $R \times C = \left(\frac{m-l}{d}\right) \times \left(\frac{m-l}{3d}\right)$. We thus obtain, for the nondecimated and decimated cases, respectively

$$\mathrm{fc}^{\mathrm{QR}}_{\mathrm{nondec}} \approx \frac{8m^3}{9} - \frac{8m^3}{81} \quad (41)$$

and

$$\mathrm{fc}^{\mathrm{QR}}_{\mathrm{dec}} \approx c\left(\frac{8(m-l)^3}{9d^3} - \frac{8(m-l)^3}{81d^3}\right). \quad (42)$$

Polynomial Root Finding Operation Counts: For a real polynomial of order κ, the approximate flop count associated with the root finding operation is given by $\mathrm{fc}^{\mathrm{root}} \approx \frac{20}{3}\kappa^3$ [14]. For complex data, the count is about a factor of 10 larger; in our case $\kappa = \left(\frac{m-l}{3d}\right)$. Therefore, we have

$$\mathrm{fc}^{\mathrm{root}}_{\mathrm{nondec}} \approx \frac{200}{81}m^3 \quad (43)$$

and

$$\mathrm{fc}^{\mathrm{root}}_{\mathrm{dec}} \approx c\left(\frac{200(m-l)^3}{81d^3}\right). \quad (44)$$

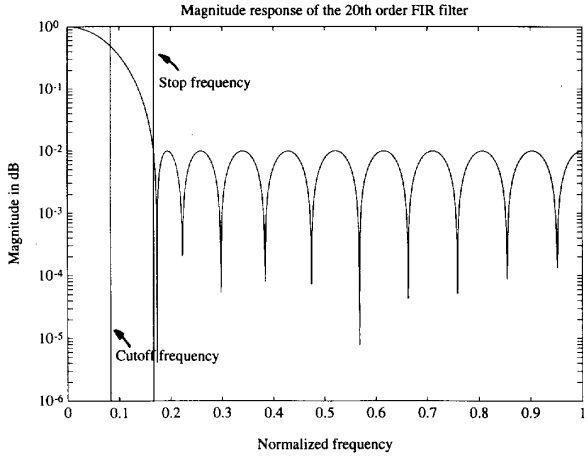


Fig. 2. Frequency response of the 20th-order equiripple FIR filter.

Filtering Operation Counts: For the decimated case, we must also include the operation count associated with the filtering operation. From (5) and (6), the approximate flop count associated with the matrix multiplication induced by this filter is thus given by

$$\mathrm{fc}^{\mathrm{filter}}_{\mathrm{dec}} \approx c\left(\frac{m-l}{d}(l+1)\right) \times 6. \quad (45)$$

The factor of 6 arises because there are six flops (four multiplies and two adds) per complex multiplication.

V. Examples and Simulation Studies

Examples using the statistical analysis results are presented here that demonstrate the advantages of using decimation. Simulations are also presented for full spectrum data sets to demonstrate the estimation ability of the modified TLS-Prony method developed in Section IV.

A. Single Undamped Mode

In this example, we assume one snapshot of data of length $m = 140$. We assume a single exponential located on the unit circle and an SNR of 5 dB. We compare estimates of this exponential using no decimation and using decimation by $d = 6$. Using the filter design procedure outlined above, we obtain an FIR filter whose frequency response is shown in Fig. 2. Here, we used $c = 2d$ overlapped estimations and an equiripple FIR filter of order $l = 20$. Note that we could decrease the order of the FIR filter to achieve the same stopband rejection, but in doing so, we will obtain a less-flat passband, which results in increased variance. Note also that in this case, we do not have out-of-band signal poles; therefore, there will be no bias in the pole estimates (to a first order).

Fig. 3 shows the theoretical variance of the estimated pole versus prediction order for various decimation factors as compared with the CRB. From this figure, we can see that the minimum variance occurs at a lower prediction order for higher d. The minimum occurs for a prediction order equal to about one third of the decimated data length (i.e., $L =$

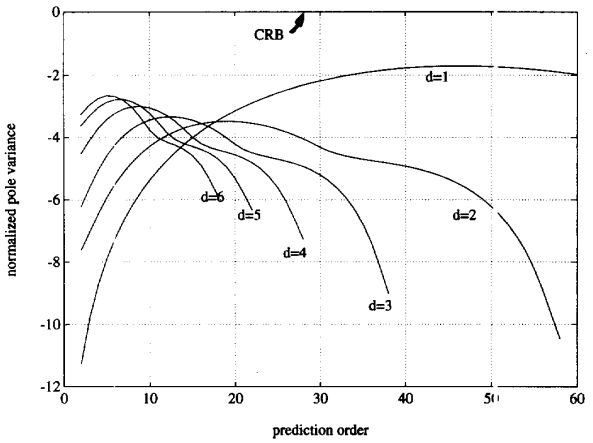


Fig. 3. $10\log_{10}(\mathrm{CRB}/\mathrm{Var}(p))$ for various d and prediction orders for a single undamped mode.

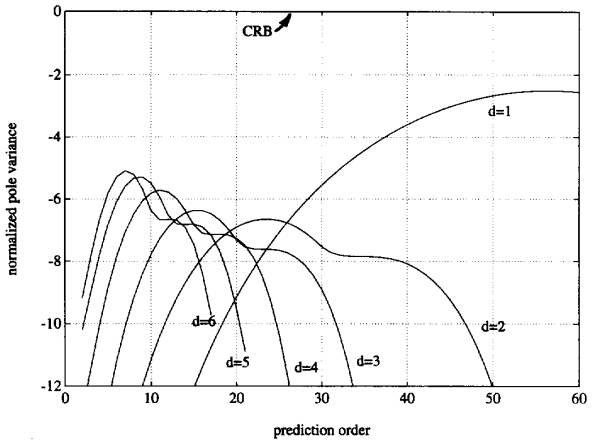


Fig. 4. $10\log_{10}(\mathrm{CRB}/\mathrm{Var}(p_1))$ for various d and prediction orders for two undamped modes.

$\frac{m'_d}{3}$), which is consistent with results for nondecimated data [15], [20]. This shows that the best performance for the decimated cases occurs at lower prediction orders than for the nondecimated case, thus reducing the computational load.

We note, however, that since the data has to be filtered prior to decimation, the curves for the decimated cases peak at about 1.5 dB lower than for the nondecimated case for various decimation factors. The performance loss is due to the fact that the transient response portion of the filter output (20 points for this case) needs to be discarded. Note that for a fixed-length FIR filter, this performance degradation becomes smaller as the data length is increased since the percent difference between the original and filtered data lengths decreases.

B. Two Undamped Modes

In a second example, we make the same assumptions above, except that there are two equal energy exponentials located on the unit circle one Fourier bin apart (i.e., $\Delta f = \frac{1}{m} = \frac{1}{140}$). The total SNR is assumed to be 8 dB in this case in order to maintain 5 dB SNR/pole. Fig. 4 shows the theoretical variance of the estimates for one of the poles (the variance of the other pole is similar). We can see that the characteristics are much the same as in the one pole case, the difference being higher variances due to the presence of each pole's neighbor.

C. Monte-Carlo Simulation of an Undamped Ten-Mode Case

We now present a set of simulations for a general ten undamped mode case. In these simulations, we have $N = 1$ snapshots and $n = 10$ poles present in the data. The amplitude coefficients all have unit magnitude; the phases of the amplitude coefficients are chosen randomly. We consider two data lengths $m = 140$ and $m = 560$ data points. Fig. 5 shows the locations of the ten poles; each is indicated by an "x." Five-hundred independent Monte-Carlo simulations are performed by adding noise to the data such that the total SNR is 20 dB (10 dB/pole). Estimates for the poles are obtained using the TLS-Prony algorithm without decimation (i.e., $d = 1$) and with decimation using a decimation factor of $d = 6$.

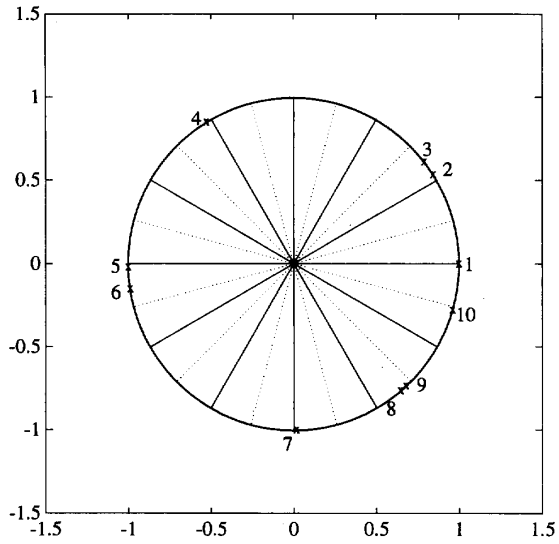


Fig. 5. True pole locations for an undamped ten-mode case.

For the decimation results, the FIR filter is the same as the one used in the previous examples, with frequency response shown in Fig. 2 (thus, $c = 12$ and $l = 20$). The prediction order is $L = \lfloor \frac{m-l}{3d} \rfloor$; the numbers of singular values retained in the simulations are 10 for $d = 1$ and the number of poles in each subestimation section for $d = 6$ ($\{2,3,2,0,1,1,2,2,0,1,3,3\}$ for this case). The prediction orders used correspond to one third of the effective data lengths in the two cases as was suggested by Examples 1 and 2. Prior to the calculation of the amplitude coefficients, poles with magnitude larger than 1.15 are eliminated to avoid poor conditioning in the least squares solution of the amplitude coefficients.

Performance Comparison: Tables I–III summarize the performance of the various methods for this example.

TABLE I
THEORETICAL AND SIMULATION VARIANCES AND MSE'S FOR THE UNDAMPED TEN POLES $m = 140$ DATA POINT CASE (ALL VALUES ARE IN DECIBELS)

pole		$d = 1$			$d = 6$			
number	CRB	Theory	Sim.	Sim. MSE	Theory	Sim.	Theory MSE	Sim. MSE
1	-63.4	-61.8	-61.5	-61.5	-60.5	-60.6	-60.3	-60.4
2	-63.0	-60.7	-60.9	-60.8	-59.7	-58.9	-59.5	-58.7
3	-63.0	-60.7	-61.0	-60.9	-60.4	-58.7	-59.9	-58.0
4	-63.6	-61.4	-61.2	-61.2	-60.9	-61.0	-60.9	-60.9
5	-63.4	-62.3	-62.3	-62.2	-60.6	-60.4	-60.1	-59.9
6	-63.4	-62.3	-62.2	-62.2	-60.7	-60.5	-60.4	-60.3
7	-63.5	-61.4	-61.1	-61.1	-61.0	-60.8	-60.9	-60.7
8	-55.1	-51.4	-51.0	-50.9	-47.3	-48.4	-47.0	-48.1
9	-55.1	-51.4	-51.3	-51.1	-46.8	-47.9	-46.5	-47.4
10	-63.4	-61.8	-61.6	-61.6	-59.6	-59.6	-59.5	-59.6

TABLE II
THEORETICAL AND SIMULATION BIASES FOR THE POLES ($m = 140$, UNIT=10^{-3})

pole number	Theory	Sim.
1	0.078-0.179i	0.088-0.196i
2	0.166-0.033i	0.255-0.037i
3	0.330+0.072i	0.449+0.188i
4	0.042+0.085i	0.005+0.100i
5	-0.082-0.314i	-0.074-0.324i
6	0.086+0.205i	0.018+0.184i
7	0.076-0.077i	0.067-0.136i
8	1.251-0.171i	0.847-0.693i
9	-1.121-0.470	-1.134-0.625i
10	-0.095+0.142i	-0.047+0.021i

We first consider Table I, which shows the $m = 140$ data point case. From Table I, we see that the Monte-Carlo variances for the $d = 1$ case were 1.2 to 4.1 dB away from their CRB's. Note that the Monte-Carlo variances are within 0.5 dB of those predicted by the theory, which substantiates the theory (the differences are due to the fact that theory is only a first-order analysis). Note also that the estimates have negligible bias, which is shown by the fact that the pole variances are very close to their MSE's.

The Monte-Carlo variances for the $d = 6$ case were 2.6 to 7.2 dB away from their CRB's. This represents an average 1.7 dB loss for the entire $d = 6$ system versus the nondecimated system. The main cause of the performance loss is the 20 data point loss due to the transient response of the FIR filter outputs. Note that the Monte-Carlo variances are within 1.7 dB of statistical theory. By comparing the simulation variances and MSE's, we see that some of pole estimates are slightly biased. When decimation is used, Theorem 4.1 gives an analytical expression for the bias; theoretical biases are compared to biases obtained from Monte-Carlo simulations with good agreement in most cases. This is shown in Table II.

TABLE III
THEORETICAL AND SIMULATION VARIANCES AND MSE'S FOR THE POLES FOR THE 560 DATA POINT CASE (ALL VALUES ARE IN DECIBELS)

pole		$d = 1$			$d = 6$			
number	CRB	Theory	Sim.	Sim. MSE	Theory	Sim.	Theory MSE	Sim. MSE
1	-81.6	-79.9	-79.6	-79.6	-79.7	-79.8	-79.7	-79.8
2	-81.6	-79.6	-79.7	-79.7	-79.4	-79.4	-79.4	-79.3
3	-81.6	-79.6	-79.7	-79.7	-79.3	-79.4	-79.3	-79.4
4	-81.7	-79.9	-79.7	-79.6	-79.7	-79.7	-78.9	-79.0
5	-81.6	-79.8	-79.8	-79.8	-79.5	-79.2	-79.4	-79.1
6	-81.6	-79.8	-79.7	-79.7	-79.4	-79.4	-79.1	-78.9
7	-81.7	-79.9	-79.9	-79.8	-79.7	-79.4	-79.6	-79.3
8	-81.6	-79.9	-80.0	-80.0	-79.6	-79.4	-79.0	-79.0
9	-81.6	-80.0	-80.0	-80.0	-79.5	-79.4	-79.5	-79.4
10	-81.7	-80.0	-80.1	-80.1	-79.3	-79.7	-79.3	-79.7

If the data length is increased to $m = 560$ points, the theoretical and simulation results show even better agreement, as is shown in Table III. In this case, the overall loss using decimation is less than 0.4 dB, compared with the nondecimated case. In addition, the simulation variances are within 0.4 dB of the theoretically derived variances, and the bias of the pole estimates is significantly reduced.

Operation Count Comparison: The $d = 1$ and $d = 6$ estimation procedures are now compared on the basis of their computational costs. We only compare the results for $m = 140$ data points. Using MATLAB, the "economical" version of the SVD operation, the left division operation (using QR decomposition to solve least squares problems) and the root-finding operations required an average of 16.5 Mflops for each of the Monte-Carlo simulations for the $d = 1$ case. Each of the 12 $d = 6$ SVD's, QR decompositions, polynomial root findings, and filtering operations required an average of 53.7 Kflops, resulting in a total of 644.4 Kflops for each Monte-Carlo simulation. The computational savings for the SVD's, QR decompositions, polynomial root findings, and filtering operations in this example using decimation was, thus, a factor of about 25.6. This compares with a savings factor of 24.0, which is predicted by (39)–(45) for this scenario. The average total flop counts (including all operations) for the nondecimated and decimated Monte-Carlo simulations were 17.4 Mflops and 698 Kflops, respectively, to give a savings factor of 24.9. Note that the four computational cost components we have detailed make up about 92% of the total computations. With higher decimation factors, the savings are even more substantial.

VI. CONCLUSION

In this paper, we have developed a TLS-Prony estimation algorithm that incorporates data decimation. We also have developed a statistical analysis for estimated poles of this algorithm. We have shown through examples using this analysis

that decimation provides a minimum variance for estimated poles that occurs at a prediction order that is smaller than the optimal prediction order for nondecimated data by a factor of d, thus allowing for computational savings. We have shown that this benefit is obtained at the expense of pole variance performance due to the filtering that is required; this expense becomes smaller for longer data lengths. We have also shown how the modified TLS-Prony method can be used on full spectrum data one band at a time to realize the computational savings in a more general signal framework.

With this decimation procedure, we are now able to make a well-quantified tradeoff of accuracy for computation when using the TLS-Prony estimation procedure.

APPENDIX
PROOF OF THEOREM

From (13), we can make the following substitutions:

$$\left(S'^0 + \widetilde{S'^0}\right)\left(b' + \tilde{b}'\right) = -\left(s'^0 + \widetilde{s'^0}\right) \tag{46}$$

where s'^0 is the noise-free version of y'^0, $\widehat{Y'^0} = S'^0 + \widetilde{S'^0}$, $\widehat{y'^0} = s'^0 + \widetilde{s'^0}$, and $\hat{b}' = b' + \tilde{b}'$. We can see that the $\tilde{}$ terms are small perturbations for the high SNR case, which is assumed. Multiplying out (46) and retaining only the first-order $\tilde{}$ terms gives[1]

$$S'^0 b' + \widetilde{S'^0} b' + S'^0 \tilde{b}' = -s'^0 - \widetilde{s'^0}. \tag{47}$$

Now note that $S'^0 b' = -s'^0$ since they are the noiseless terms. Equation (47) thus becomes

$$S'^0 \tilde{b}' = -\left(\widetilde{s'^0} + \widetilde{S'^0} b'\right). \tag{48}$$

Multiplying both sides through by $S'^0 S'^{0+}$ and noting that $S'^0 S'^{0+} S'^0 = S'^0$, we obtain

$$S'^0 \tilde{b}' = -S'^0 S'^{0+}\left(\widetilde{s'^0} + \widetilde{S'^0} b'\right). \tag{49}$$

[1] Note that the approximation is valid since the matrices $[\widehat{y'^0} \;:\; \widehat{Y'^0}]$ and $[s'^0 \;:\; S'^0]$ have the same rank.

Let $Y'^0 = S'^0 + W'^0$ and $y'^0 = s'^0 + w'^0$, where W'^0 and w'^0 are the appropriate noise matrices (i.e., of the form given in (13) and composed of the noise sequences in (12)). Thus, we can see that $\widetilde{S'^0}$ and $\widetilde{s'^0}$ are perturbations caused by W'^0, w'^0, and the SVD truncation. By using perturbation analysis [21] on the matrices $[s'^0 \; S'^0]$, $[y'^0 \; Y'^0]$, and $\left[\widehat{y'^0} \; \widehat{Y'^0}\right]$, it can be shown that to first-order approximation

$$S'^{0+}\left[\widetilde{s'^0} \quad \widetilde{S'^0}\right] = S'^{0+}[w'^0 \quad W'^0]. \tag{50}$$

Thus, (49) can be written as

$$S'^0 \tilde{b}' = -S'^0 S'^{0+} \epsilon \tag{51}$$

where $\epsilon = w'^0 + W'^0 b'$.

Observing the data model and the formulation of the S'^0 matrix, we can write S'^0 as (52), which appears at the bottom of the page, or

$$S'^0 = HG'. \tag{53}$$

Equation (51) thus becomes

$$HG'\tilde{b}' = -HG'S'^{0+}\epsilon. \tag{54}$$

Now, note by definition that the true and estimated Lth-order characteristic polynomials are $B(z) = 1 + b'_1 z + b'_2 z^2 + \cdots + b'_L z^L$ and $\widehat{B'}(z) = 1 + \hat{b}'_1 z + \hat{b}'_2 z^2 + \cdots + \hat{b}'_L z^L$, respectively. Hence, $B(p'_i) = 0$, and $\widehat{B'}$ and $(\hat{p}'_i) = 0$.

We can use a first-order Taylor expansion to find an expression for the error in the estimated pole locations. For each $\hat{p}'_i$, we obtain (55), which apears at the top of the next page, or, to first order

$$\left(\hat{p}'_i - p'_i\right) = -\frac{1}{\eta'_i}[p'_i \quad p'^2_i \quad \cdots \quad p'^L_i]\tilde{b}'. \tag{56}$$

Thus, for all of the n' true poles, we obtain

$$\widehat{P'} - P' = -F'G'\tilde{b}'. \tag{57}$$

Since H is full rank (this can be seen by noting that each block of rows is simply a Vandermonde matrix derived from

$$S'^0 = \begin{bmatrix} x_1(1) & x_2(1) & \cdots & x_{n'}(1) \\ x_1(1)p_1^d & x_2(1)p_2^d & \cdots & x_{n'}(1)p_{n'}^d \\ \vdots & \vdots & & \vdots \\ x_1(1)p_1^{m-(L+1)d} & x_2(1)p_2^{m-(L+1)d} & \cdots & x_{n'}(1)p_{n'}^{m-(L+1)d} \\ x_1(2) & x_2(2) & \cdots & x_{n'}(2) \\ x_1(2)p_1^d & x_2(2)p_2^d & \cdots & x_{n'}(2)p_{n'}^d \\ \vdots & \vdots & & \vdots \\ x_1(2)p_1^{m-(L+1)d} & x_2(2)p_2^{m-(L+1)d} & \cdots & x_{n'}(2)p_{n'}^{m-(L+1)d} \\ & \vdots & & \\ x_1(N) & x_2(N) & \cdots & x_{n'}(N) \\ x_1(N)p_1^d & x_2(N)p_2^d & \cdots & x_{n'}(N)p_{n'}^d \\ \vdots & \vdots & & \vdots \\ x_1(N)p_1^{m-(L+1)d} & x_2(N)p_2^{m-(L+1)d} & \cdots & x_{n'}(N)p_{n'}^{m-(L+1)d} \end{bmatrix} G' \tag{52}$$

$$
\begin{aligned}
0 &= \widehat{B'}\left(\widehat{p'_i}\right) \\
&= \widehat{B'}(p'_i) + \frac{\partial}{\partial z}\widehat{B'}(z)|_{z=p'_i}\left(\widehat{p'_i} - p'_i\right) + \text{(higher order terms)} \\
&= \widehat{B'}(p'_i) - B'(p'_i) + \frac{\partial}{\partial z}\widehat{B'}(z)|_{z=p'_i}\left(\widehat{p'_i} - p'_i\right) + \text{(higher order terms)} \\
&\approx 1 + \widehat{b'_1}p'_i + \widehat{b'_2}p_i'^2 + \cdots + \widehat{b'_L}p_i'^L - \left(1 + b'_1p'_i + b'_2p_i'^2 + \cdots + b'_Lp_i'^L\right) \\
&+ \left(\widehat{b'_1} + 2\widehat{b'_2}p'_i + \cdots + L\widehat{b'_L}p_i'^{(L-1)}\right)\left(\widehat{p'_i} - p'_i\right) \\
&\approx [p'_i \quad p_i'^2 \quad \cdots \quad p_i'^L]\begin{bmatrix} \widehat{b'_1} - b'_1 \\ \widehat{b'_2} - b'_2 \\ \vdots \\ \widehat{b'_L} - b'_L \end{bmatrix} + [b'_1 \quad b'_2 \quad \cdots \quad b'_L]\begin{bmatrix} 1 \\ 2p'_i \\ \vdots \\ Lp_i'^{(L-1)} \end{bmatrix}\left(\widehat{p'_i} - p'_i\right) \\
&= [p'_i \quad p_i'^2 \quad \cdots \quad p_i'^L]\left(\widehat{b'} - b'\right) + \eta'_i\left(\widehat{p'_i} - p'_i\right)
\end{aligned} \tag{55}
$$

distinct poles times a diagonal matrix of the nonzero amplitude coefficients), we can multiply (54) by $(H^*H)^{-1}H^*$ to get

$$G'\tilde{b'} = -G'S'^{0+}\epsilon \tag{58}$$

and by substituting (58) into (57), we obtain

$$\widehat{P'} - P' = F'G'S'^{0+}\epsilon. \tag{59}$$

We now note that $[w'^0 \quad W'^0]\left[\begin{smallmatrix}1\\b'\end{smallmatrix}\right]$ can be written as

$$
\begin{aligned}
[w'^0 \quad W'^0]\begin{bmatrix}1\\b'\end{bmatrix} &= \begin{bmatrix} B'e'^0(1) \\ B'e'^0(2) \\ \vdots \\ B'e'^0(N) \end{bmatrix} \\
&= \begin{bmatrix} B'K^0(s^r(1) + e(1)) \\ B'K^0(s^r(2) + e(2)) \\ \vdots \\ B'K^0(s^r(N) + e(N)) \end{bmatrix}
\end{aligned} \tag{60}
$$

where

$$
\begin{aligned}
&s^r(t) \\
&= \left[\sum_{i=n'+1}^{n} x_i(t) \quad \sum_{i=n'+1}^{n} x_i(t)p_i \ldots \sum_{i=n'+1}^{n} x_i(t)p_i^{m-1}\right]^T,
\end{aligned} \tag{61}
$$

and where B' and K^0 are given by (28) and (29). Recall that $\{e(t)\}$ are zero mean Gaussian, and thus, ϵ is multivariate Gaussian with mean

$$E[\epsilon] = P'0_\epsilon \tag{62}$$

where $P'0_\epsilon$ is in the form of (60), without the $e(t)$'s, and covariance matrix

$$
\begin{aligned}
\text{Cov}(\epsilon) &= E\left[\left([w'^0 \quad W'^0]\begin{bmatrix}1\\b'\end{bmatrix}\right)\left([w'^0 \quad W'^0]\begin{bmatrix}1\\b'\end{bmatrix}\right)^*\right] \\
&= \sigma\Sigma_\epsilon^0
\end{aligned} \tag{63}
$$

where Σ_ϵ^0 is given by (27).

Equations (59), (62), and (63) imply that the mean and covariance matrix of $\widehat{P'}$ are given by (26).

REFERENCES

[1] M. A. Rahman and K.-B. Yu, "Total least squares approach for frequency estamatin using linear prediction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, no. 10, pp. 1440–1454, Oct. 1987.

[2] W. M. Steedly and R. L. Moses, "High resolution exponential modeling of fully polarized radar returns," *IEEE Trans. Aerospace .Electron. Syst.*, vol. AES-27, no. 3, pp. 459–469, May 1991.

[3] B. Liu and F. Mintzer, "Caculation of narrow-band spectra by direct decimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, no. 6, pp. 529–534, Dec. 1978.

[4] M. P. Quirk and B. Liu, "Improving resolution for autoregressive spectral estimation by decimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, no. 3, pp. 630–637, June 1983.

[5] M. J. Villalba and B. K. Walker, "Spectrum manipulation for improved resolution," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 6, pp. 820–831, June 1989.

[6] C. J. Ying, R. L. Moses, and R. L. Dilsavor, "Perturbation analysis for pole estimates of damped exponential signals," Tech. Rep., Ohio State Univ., Dept. of Elect. Eng., ElectroSci. Lab., Aug. 1991.

[7] G. Xu, S. D. Silverstein, R. H. Roy, and T. Kailath, "Parallel implementation and performance analysis of beamspace ESPRIT," in *Proc. Int. Conf. Acoust., Speech, Signal Processing* (Toronto), May 14–17, 1991, pp. 1497–1500.

[8] M. D. Zoltowski, G. M. Kautz, and S. D. Silverstein, "Development, performance analysis, and experimental evaluation of beamspace root-MUSIC," in *Proc. Int. Conf. Acoust., Speech, Signal Processing* (Toronto), May 14–17, 1991, pp. 3049–3052.

[9] M. Zoltowski, G. Kautz, and S. Silverstein, "Beamspace root-MUSIC," *IEEE Trans. Signal Processing*, vol. 41, no. 1, pp. 344–364, Jan. 1993.

[10] S. D. Silverstein, W. E. Engeler, and J. A. Taridif, "Parallel architectures for multirate superresolution spectrum analyzers," *IEEE Trans. Circuits Syst.*, vol. 38, no. 4, pp. 449–453, Apr. 1991.

[11] H. Lee and M. Wengrovitz, "Improved high-resolution direction-finding through use of homogeneous constraints," in *Proc. Fourth ASSP Workshop Spect. Est. Modeling*, 1988, pp. 152–157.

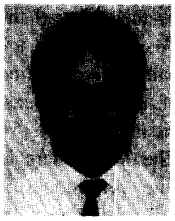
[12] P. Vaidynathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[13] T. Abatzoglou, J. Mendel, and G. Harada, "The constrained total least squares technique and its applications to harmonic superresolution," *IEEE Trans. Signal Processing*, vol. 39, no. 5, pp. 1070–1087, May 1991.

[14] G. H. Golub and C. F. VanLoan, *Matrix Computations*. Baltimore, MD: Johns Hopkins, 1989, 2nd ed.

[15] W. M. Steedly, C. J. Ying, and R. L. Moses, "Statistical analysis of TLS-based Prony techniques," *Automatica (Special Issue on Statistical Signal Processing and Control)*, vol. 30, no. 1, pp. 115–129, Jan. 1994.

[16] W. M. Steedly and R. L. Moses, "The Cramér-Rao bound for pole and amplitude estimates of damped exponential signals in noise," in *Proc. Int. Conf. Acoust., Speech, Signal Processing* (Toronto), May 14–17, 1991, pp. 3569–3572.
[17] X. Xu and K. Buckley, "Statistical performance comparison of MUSIC in element-space and bean-space," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, May 1989, pp. 2124–2127.
[18] S. Anderson, "Optimal dimension reduction for sensor array signal processing," in *Conf. Rec. 25th Asilomar IEEE Conf. Signals, Syst., Comput.*. Nov. 4–6, 1991, pp. 918–922.
[19] P. Stoica and A. Nehorai, "Comparative performance study of element-space and beam-space MUSIC estimators," *Circuits, Syst. Signal Processing*, vol. 10, pp. 285–292, 1991.
[20] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-38, no. 5, pp. 814–824, May 1990.
[21] ——, "A perturbation property of the TLS-LP method," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 11, pp. 2004–2005, Nov. 1990.



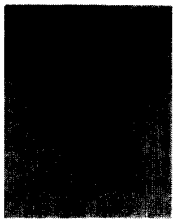
William M. Steedly (S'86–M'93) received the B.S. degree in electrical engineering from Virginia Polytechnic Institute and State University, Blacksburg, in 1988 and the M.S. and Ph.D. degrees in electrical engineering from The Ohio State University, Columbus, in 1989 and 1992, respectively.

During the 1988–1989 academic year, he was an Ohio State University Fellow. During the 1989–1992 academic years, he was an Air Force Laboratory Graduate Fellow in the Department of Electrical Engineering of The Ohio State University. He spent the summer of 1990 at Wright Labs, Wright Patterson AFB, OH. He is currently a member of the Technical Staff of The Analytic Sciences Corporation (TASC), Reston, VA. His primary research interests are in digital signal processing, including parametric modeling techniques and their application to radar signal processing.

Dr. Steedly is a member of Eta Kappa Nu, Tau Beta Pi, Phi Kappa Phi, and Sigma Xi

Ching-Hui J. Ying (S'91) was born in Keelung, Taiwan, on August 3, 1967. He received the diploma in electronic engineering from National Taipei Institute of Techology, Taiwan, in 1987 and the M.S. degree in electrical engineering from The Ohio State University (OSU), Columbus, in 1992, respectively. He is currently pursuing the Ph.D. degree in electrical engineering at OSU.

Since 1991, he has worked as a Graduate Research Associate with The Ohio State University ElectroScience Laboratory. His current research involves parametric modeling and neural networks. Other research interests include time series analysis, model order determination, perturbation theory, and system identification.



Randolph L. Moses (S'78–M'85–SM'90) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Virginia Polytechnic Institute and State University in 1979, 1980, and 1984, respectively.

During the summer of 1983, he was a SCEEE Summer Faculty Research Fellow at Rome Air Development Center, Rome, NY. From 1984 to 1985, he was with the Eindhoven University of Technology, Eindhoven, The Netherlands, as a NATO Post-doctoral Fellow. Since 1985, he has been with the Department of Electrical Engineering, The Ohio State University, and is currently an Associate Professor there. His research interests are in digital signal processing and include parametric time series analysis, radar signal processing, system identification, and model reduction.

Dr. Moses is a member of Eta Kappa Nu, Tau Beta Pi, Phi Kappa Phi, and Sigma Xi.